

STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS



ON IDENTIFICATION OF BAYESIAN DSGE MODELS*

BY

GARY KOOP, M. HASHEM PESARAN AND RON P. SMITH

No. 11-08

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

On Identification of Bayesian DSGE Models*

Gary Koop, University of Strathclyde
M. Hashem Pesaran, Cambridge University, and USC
Ron P. Smith, Birkbeck College

March 28, 2011

Abstract

In recent years there has been increasing concern about the identification of parameters in dynamic stochastic general equilibrium (DSGE) models. Given the structure of DSGE models it may be difficult to determine whether a parameter is identified. For the researcher using Bayesian methods, a lack of identification may not be evident since the posterior of a parameter of interest may differ from its prior even if the parameter is unidentified. We show that this can even be the case even if the priors assumed on the structural parameters are independent. We suggest two Bayesian identification indicators that do not suffer from this difficulty and are relatively easy to compute. The first applies to DSGE models where the parameters can be partitioned into those that are known to be identified and the rest where it is not known whether they are identified. In such cases the marginal posterior of an unidentified parameter will equal the posterior expectation of the prior for that parameter conditional on the identified parameters. The second indicator is more generally applicable and considers the rate at which the posterior precision gets updated as the sample size (T) is increased. For identified parameters the posterior precision rises with T , whilst for an unidentified parameter its posterior precision may be updated but its rate of update will be slower than T . This result assumes that the identified parameters are \sqrt{T} -consistent, but similar differential rates of updates for identified and unidentified parameters can be established in the case of super consistent estimators. These results are illustrated by means of simple DSGE models.

JEL Classifications: C11, C15, E17

Key Words: Bayesian identification, DSGE models, posterior updating, New Keynesian Phillips Curve.

*We are grateful to Cheng Hsiao, Dale Poirier, and Ivan Jeliazkov for helpful discussions.

1 Introduction

Soon after rational expectations (RE) models were widely adopted in economics there was concern about the issue of observational equivalence (Sargent, 1976, McCallum, 1979) and the identification of the parameters of the RE models (Wallis, 1980, Pesaran, 1981, 1987, Pudney, 1982). Observational equivalence concerns whether one can distinguish different models, such as RE and non-RE models; the closely related issue of identification concerns the conditions under which it is possible to estimate the parameters of a particular model from available data. During the 1990s interest in identification waned, partly because of the shift in focus to calibration, where it is assumed that the parameters are known *a priori*, perhaps from microeconomic evidence.¹ Kydland and Prescott (1996) argue that the task of computational experiments of the sort they conduct is to derive the quantitative implications of the theory rather than to measure economic parameters, one of the primary objects of econometric analysis.

Over the past ten years it has become more common to estimate, rather than calibrate, dynamic stochastic general equilibrium (DSGE) models, often using Bayesian techniques (see, among many others, DeJong, Ingram and Whiteman, 2000, Smets and Wouters, 2003, 2007 and An and Schorfheide, 2007). In this context the issue of identification has attracted renewed attention. Questions have been raised about the identification of particular equations of the standard new Keynesian DSGE model, such as the Phillips curve (Mavroeidis, 2005, Nason and Smith, 2008, Kleibergen and Mavroeidis, 2009, Dees et al., 2009, and others), or the Taylor rule, Cochrane (2007). There have also been questions about the identification of DSGE systems as a whole. Canova and Sala (2009) conclude: "it appears that a large class of popular DSGE structures are only very weakly identified". Iskrev (2010b) concludes "the results indicate that the parameters of the Smets and Wouters (2007) model are quite poorly identified in most of the parameter space". Other recent papers which consider determining the identification of DSGE systems are Andrieu (2010), Iskrev (2010a), Komunjer and Ng (2010), who provide rank and order conditions for local identification based on the spectral density matrix, and Muller (2010), who suggests measures of prior sensitivity and prior informativeness based on the derivative of the posterior mean with respect to a particular parameterization of the prior mean.

The 1980s literature on the identification of RE models tended to assume that the system included observed exogenous variables, whereas most current DSGE systems do not contain such variables. While most of the DSGE literature has focussed on the regular or determinate case where there is a unique solution to the linear RE system, there has been some interest in the indeterminate case, where there are multiple solutions (e.g. Clarida, Gali, and Gertler, 2000, Beyer and Farmer, 2004 and Lubik and Schorfheide, 2004). The indeterminate case also raises interesting identification issues.

¹The calibrators' practice of basing the estimates of the structural parameters of macro models on macroeconomic evidence has been criticized by microeconomists, such as Hansen and Heckman (1996).

Unlike the simpler simultaneous equations model (SEM) the non-linear nature of the cross equation restrictions in DSGE models makes it often very difficult to analytically check identification. The RE structure means that they require more identifying restrictions than comparable SEMs. Although the approximate solution of DSGE models is taken to be linear, the structural parameters are complicated non-linear functions of the parameters of the linearized (reduced form) model and as a result the likelihood function for the structural parameters may be very badly behaved. This has led many to look at other features than the likelihood, such as impulse response functions or impact effects. The form of the likelihood may also cause problems for understanding certain features of the posterior, e.g. Herbst (2010). When the model involves unobserved variables the solution is of a VARMA form rather than a VAR. Thus some of the associated reduced form parameters may not be identified. The requirement for a determinate solution also puts restrictions on the joint parameter space, which may create dependence between identified and unidentified parameters.

Faced with these difficulties, it is common practice in Bayesian DSGE modelling to compare posteriors to priors as informal indicators of identification. We discuss how this can be misleading, since, as we show, priors can differ from posteriors even for unidentified parameters. We then propose two different Bayesian indicators of identification that do not suffer from this drawback. The first draws on results from Poirier (1998) and concerns the case where the parameters can be partitioned into those known to be identified and those where it is uncertain whether they are identified. Then the marginal posterior of an unidentified parameter will equal the posterior expectation of the prior for that parameter conditional on the identified parameters. The marginal posterior and posterior expectation of the prior can be computed as a by-product of estimating a DSGE model (e.g. using MCMC methods) and compared. However, this indicator relies on the assumption that parameters can be separated into those which the researcher knows are identified and those for which identification is uncertain. Unfortunately, as we show in this paper, for the researcher working with the structural parameters of DSGE models, this assumption may not hold. Hence, we propose a second Bayesian indicator of local identification. This uses the fact that whilst for identified parameters the posterior precision rises with T , for an unidentified parameter its posterior precision may be updated but its rate of update will be slower than T . This suggests a strategy where the researcher simulates larger and larger data sets and observes the behavior of the posterior as sample size increases. Empirical illustrations show the usefulness of both these approaches for checking for the presence and strength of identification.

The paper is organized as follows. Section 2 discusses the theory of rational expectations DSGE models and provides some simple theoretical examples. Section 3 discusses the econometrics. It is broken into sub-sections on i) general identification issues; ii) existing Bayesian approaches to identification in DSGE models; iii) how (and when) these existing approaches can be used to check for presence and/or strength of identification; and iv) asymptotic results which show the behavior of the posterior for non-identified parameters in large sam-

ples and how these can be used to check for identification. Section 4 provides several empirical illustrations of the methods developed or discussed in Section 3. Section 5 concludes.

2 The Theory of Rational Expectations DSGE Models

2.1 A General Framework

Most macroeconomic DSGE models are constructed by linearizing an underlying non-linear model around its steady state, where θ is a vector of deeper parameters of this underlying model.² Consider a linearized rational expectations model for an $n \times 1$ vector of stationary variables of interest, \mathbf{y}_t , $t = 1, 2, \dots, T$. These would usually be measured as deviations from their steady states. Denote expectations as $E_t(\mathbf{y}_{t+1}) = E(\mathbf{y}_{t+1} | \mathcal{I}_t)$ where \mathcal{I}_t is the information set available at time t .³ There is also a $k \times 1$ vector of observed exogenous variables⁴ \mathbf{x}_t and an $n \times 1$ vector of unobserved variables \mathbf{u}_t . We assume that both the exogenous and unobserved variables follow VAR(1) processes without feedbacks. Then the system can be written

$$\begin{aligned} \mathbf{A}_0(\theta)\mathbf{y}_t &= \mathbf{A}_1(\theta)E_t(\mathbf{y}_{t+1}) + \mathbf{A}_2(\theta)\mathbf{y}_{t-1} + \mathbf{A}_3(\theta)\mathbf{x}_t + \mathbf{u}_t, \\ \mathbf{x}_t &= \Phi_x\mathbf{x}_{t-1} + \mathbf{v}_t, \quad \mathbf{u}_t = \Phi_u\mathbf{u}_{t-1} + \varepsilon_t, \end{aligned} \quad (1)$$

where θ are the structural parameters. We treat the VAR(1) parameters for \mathbf{x}_t or \mathbf{u}_t , Φ_i , $i = x, u$, as not being specified by macroeconomic theory. This structure assumes that there is no feedback from \mathbf{y}_t to \mathbf{x}_t or \mathbf{u}_t . ε_t is a vector of mean zero, serially uncorrelated, structural shocks, with $E(\varepsilon_t\varepsilon_t') = \Omega(\theta)$. It is common in the literature to assume that $\Omega(\theta) = \mathbf{I}_n$. For Bayesian or maximum likelihood estimation, ε_t is typically assumed to be normally distributed. Notice that if $\Phi_u = \mathbf{0}$, the structural shocks enter the equations directly.

If $\mathbf{A}_0(\theta)$ is nonsingular, then (1) can be written

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}_0(\theta)^{-1}\mathbf{A}_1(\theta)E_t(\mathbf{y}_{t+1}) + \mathbf{A}_0(\theta)^{-1}\mathbf{A}_2(\theta)\mathbf{y}_{t-1} \\ &\quad + \mathbf{A}_0(\theta)^{-1}\mathbf{A}_3(\theta)\mathbf{x}_t + \mathbf{A}_0(\theta)^{-1}\mathbf{u}_t. \end{aligned} \quad (2)$$

The solution of such systems is discussed in Binder and Pesaran (1995, 1997) and Sims (2002). The solution method proposed by Binder and Pesaran involves finding an $n \times n$ matrix $\mathbf{C}(\theta)$ such that in terms of the quasi-difference transformation $\mathbf{Y}_t = \mathbf{y}_t - \mathbf{C}(\theta)\mathbf{y}_{t-1}$, the model only involves future expectations,

²We focus on the linearized case where the estimated parameters are non-linear functions of the structural parameters. Our analysis should continue to be applicable were an exact solution to be available, since again there would be a non-linear relation between the estimated and structural parameters.

³In some papers in the literature, expectations are taken using information at time $t-1$.

⁴Many DSGE models are closed, without exogenous variables, but we include them for completeness.

where $\mathbf{C}(\boldsymbol{\theta})$ is a solution of the following quadratic matrix equation

$$\mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})^2 - \mathbf{A}_0(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta}) + \mathbf{A}_2(\boldsymbol{\theta}) = \mathbf{0}. \quad (3)$$

Then assuming $[\mathbf{I}_n - \mathbf{A}_0(\boldsymbol{\theta})^{-1}\mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]$ is non-singular, we obtain

$$\mathbf{Y}_t = \mathbf{F}(\boldsymbol{\theta})E_t(\mathbf{Y}_{t+1}) + \mathbf{W}_t,$$

where

$$\begin{aligned} \mathbf{F}(\boldsymbol{\theta}) &= [\mathbf{A}_0(\boldsymbol{\theta}) - \mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1} \mathbf{A}_1(\boldsymbol{\theta}), \\ \mathbf{W}_t &= [\mathbf{A}_0(\boldsymbol{\theta}) - \mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1} [\mathbf{A}_3(\boldsymbol{\theta})\mathbf{x}_t + \mathbf{u}_t]. \end{aligned}$$

There will be a unique solution if there exists a real matrix solution to (3) such that all the eigenvalues of $\mathbf{C}(\boldsymbol{\theta})$ lie inside or on the unit circle, and all the eigenvalues of $\mathbf{F}(\boldsymbol{\theta})$ lie strictly inside the unit circle. In such cases the unique solution is given by

$$\mathbf{y}_t = \mathbf{C}(\boldsymbol{\theta})\mathbf{y}_{t-1} + \sum_{h=0}^{\infty} \mathbf{F}(\boldsymbol{\theta})^h E_t(\mathbf{W}_{t+h}).$$

When, as assumed above, \mathbf{x}_t and \mathbf{u}_t follow a VAR(1) process, and the roots of the $\boldsymbol{\Phi}_i$, $i = x, u$ lie on or inside the unit circle, then

$$\begin{aligned} E_t(\mathbf{x}_{t+h}) &= \boldsymbol{\Phi}_x^h \mathbf{x}_t \\ E_t(\mathbf{u}_{t+h}) &= \boldsymbol{\Phi}_u^h \mathbf{u}_t \end{aligned}$$

and since \mathbf{x}_t and \mathbf{u}_t are independent

$$E_t(\mathbf{W}_{t+h}) = [\mathbf{A}_0(\boldsymbol{\theta}) - \mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1} [\mathbf{A}_3(\boldsymbol{\theta})\boldsymbol{\Phi}_x^h \mathbf{x}_t + \boldsymbol{\Phi}_u^h \mathbf{u}_t]$$

therefore the solution has the form

$$\begin{aligned} \mathbf{y}_t &= \mathbf{C}(\boldsymbol{\theta})\mathbf{y}_{t-1} + \\ &\quad \sum_{h=0}^{\infty} \mathbf{F}^h(\boldsymbol{\theta}) [\mathbf{A}_0(\boldsymbol{\theta}) - \mathbf{A}_1(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1} [\mathbf{A}_3(\boldsymbol{\theta})\boldsymbol{\Phi}_x^h \mathbf{x}_t + \boldsymbol{\Phi}_u^h \mathbf{u}_t], \end{aligned} \quad (4)$$

which we can write

$$\mathbf{y}_t = \mathbf{C}(\boldsymbol{\theta})\mathbf{y}_{t-1} + \mathbf{G}_1(\boldsymbol{\theta}, \boldsymbol{\phi}_x)\mathbf{x}_t + \mathbf{G}_2(\boldsymbol{\theta}, \boldsymbol{\phi}_u)\mathbf{u}_t. \quad (5)$$

where $\boldsymbol{\phi}_i = \text{vec}(\boldsymbol{\Phi}_i)$, $i = x, u$. The matrices $\mathbf{G}_i(\boldsymbol{\theta}, \boldsymbol{\phi}_i)$ $i = x, u$, can be obtained using the method of undetermined coefficients (see Blinder and Pesaran, 1997, for details). Notice that the coefficient matrix for the lagged dependent variable vector is just a function of $\boldsymbol{\theta}$, and not $\boldsymbol{\phi}_x$ or $\boldsymbol{\phi}_u$.

Likelihood-based estimation of this model is straightforward. If $\boldsymbol{\Phi}_u = 0$, this is just a VAR with exogenous variables and the likelihood function is easily

obtained. In general where the unobserved components of the model are serially correlated, the rational expectations solution will involve moving average components and it is more convenient to write the model as a state space model where Kalman filtering techniques can be used to evaluate the likelihood function. In such cases a simple analytical relationship between the structural and reduced form parameters might not be available, which further complicates the analysis of identification of the structural parameters. In the next sub-sections, we use some simple special cases of DSGE models where the RE solution is available analytically to clearly show how identification issues arise. For notational simplicity, we do not make the dependence on θ explicit.

2.2 DSGE models without lags

Abstracting from lagged values and exogenous regressors (1) simplifies to

$$\begin{aligned}\mathbf{A}_0 \mathbf{y}_t &= \mathbf{A}_1 E_t(\mathbf{y}_{t+1}) + \boldsymbol{\varepsilon}_t, \\ E(\boldsymbol{\varepsilon}_t) &= 0, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega},\end{aligned}\tag{6}$$

where $\boldsymbol{\varepsilon}_t$ are serially uncorrelated. If \mathbf{A}_0 is non-singular using (6) we have

$$\mathbf{y}_t = \mathbf{A}_0^{-1} \mathbf{A}_1 E_t(\mathbf{y}_{t+1}) + \mathbf{A}_0^{-1} \boldsymbol{\varepsilon}_t = \mathbf{Q} E_t(\mathbf{y}_{t+1}) + \mathbf{A}_0^{-1} \boldsymbol{\varepsilon}_t.\tag{7}$$

The regular case, where there is a unique stationary solution, arises if the non-zero eigenvalues of \mathbf{Q} lie within the unit circle. In this case, the unique solution of the model is given by

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{Q}^j \mathbf{A}_0^{-1} E_t(\boldsymbol{\varepsilon}_{t+j}).\tag{8}$$

Hence, $E_t(\mathbf{y}_{t+1}) = 0$ and the unique RE solution is given by

$$\mathbf{A}_0 \mathbf{y}_t = \boldsymbol{\varepsilon}_t.\tag{9}$$

and

$$\begin{aligned}\mathbf{y}_t &= \mathbf{A}_0^{-1} \boldsymbol{\varepsilon}_t = \mathbf{u}_t \\ E(\mathbf{u}_t \mathbf{u}_t') &= \boldsymbol{\Sigma} = \mathbf{A}_0^{-1} \boldsymbol{\Omega} \mathbf{A}_0^{-1'}.\end{aligned}\tag{10}$$

Notice that (10) provides us with a likelihood function which does not depend on \mathbf{A}_1 and, therefore, the parameters that are unique to \mathbf{A}_1 (i.e. the coefficients on the forward variables) are not identified. Furthermore, the RE model is observationally equivalent to a model without forward variables which takes the form of (9). Since what can be estimated from the data, $\boldsymbol{\Sigma}$, is not a function of \mathbf{A}_1 , all possible values of \mathbf{A}_1 are observationally equivalent in the sense that they lead to the same observed data covariance matrix. Although the coefficients in the forward solution (8) are functions of \mathbf{A}_1 , this does not identify them because $E_t(\boldsymbol{\varepsilon}_{t+j}) = 0$. Elements of \mathbf{A}_1 could be identified by certain sorts of *a priori*

restrictions, but these are likely to be rather special, rather limited in number and cannot be tested.

If the parameters of the DSGE model were thought to be known *a priori* from calibration, there would be no identification problem and the structural errors ε_{it} could be recovered and used, for instance, in calculating impulse response functions, IRFs. However, suppose someone else believed that the true model was just a set of random errors $\mathbf{y}_t = \mathbf{u}_t$, with different IRFs. There is no information in the data that a proponent of the DSGE could use to persuade the other person that the DSGE model was correct relative to the random error model. This is exactly the same point that Sargent (1976) made with respect to "natural and unnatural rate theories".

Given data, one can estimate the $n(n+1)/2$ independent elements of Σ , in (10) and the solution to this model is exactly the same as the reduced form in the classical simultaneous equations model (SEM). The familiar order condition for the SEM is that identification of \mathbf{A}_0 and Ω requires that there are n^2 *a priori* restrictions on the parameters.⁵ In contrast, the order condition for the RE model (6) requires $2n^2$ *a priori* restrictions on the parameters. As in the SEM case, it may well be that although the order condition is satisfied, the rank condition fails. In this case, this is likely since, given the structure of the model, $E(\mathbf{y}_{t+1}) = 0$, any deviation must be random. Thus, there is no variation in expected future values, which could identify their effects. Notice that this would also be true, if there were no simultaneity and \mathbf{A}_0 was an identity matrix. In this case, any stable model with rational expectations has a corresponding solution without expectations.

The above result generalizes to higher order RE models. Consider for example the model

$$\mathbf{A}_0 \mathbf{y}_t = \sum_{i=1}^p \mathbf{A}_i E_t(\mathbf{y}_{t+i}) + \varepsilon_t.$$

Once again the unique stable solution of this model is also given by $\mathbf{A}_0 \mathbf{y}_t = \varepsilon_t$.

2.2.1 Example 1. A New Keynesian (NK) system without lags

As an illustration consider a standard three equation NK-DSGE model used in Benati (2010), for example, that involves only current and future variables:

$$R_t = \psi \pi_t + \varepsilon_{1t}, \quad (11)$$

$$x_t = E_t(x_{t+1}) - \sigma(R_t - E_t(\pi_{t+1})) + \varepsilon_{2t}, \quad (12)$$

$$\pi_t = \beta E_t(\pi_{t+1}) + \gamma x_t + \varepsilon_{3t}. \quad (13)$$

where $E_t(x_{t+1}) = E(x_{t+1} \mid \mathcal{J}_t)$. The model contains a monetary policy rule determining the interest rate, R_t , an IS curve determining the output gap, x_t ,

⁵ A sufficient set is that after suitable ordering \mathbf{A}_0 is lower triangular, which provides $n(n-1)/2$ restrictions, Ω is diagonal, which provides a further $n(n-1)/2$ restrictions, which together with n normalisation conditions, allows one to estimate the $n(n-1)/2$ free elements of \mathbf{A}_0 and the n diagonal elements of Ω . The plausibility of the assumptions required for such a recursive system has been widely questioned.

and a Phillips Curve determining inflation, π_t , all measured as deviations from their steady states. The errors, which are assumed to be white noise, are a monetary policy shock, ε_{1t} , a demand shock, ε_{2t} , and a supply or cost shock, ε_{3t} . These are assumed to be orthogonal. The discount factor is β and σ is the inter-temporal elasticity of substitution. This is a highly restricted system with many parameters set to zero *a priori*. For instance, output does not appear in the monetary policy rule and the coefficient of future output is exactly equal to unity in the IS equation. In terms of (6), $\mathbf{y}_t = (R_t, x_t, \pi_t)'$ and

$$\mathbf{A}_0 = \begin{pmatrix} 1 & 0 & -\psi \\ \sigma & 1 & 0 \\ 0 & -\gamma & 1 \end{pmatrix}, \quad \mathbf{A}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \sigma \\ 0 & 0 & \beta \end{pmatrix}.$$

Hence

$$\begin{aligned} \mathbf{A}_0^{-1} &= \frac{1}{\gamma\sigma\psi + 1} \begin{pmatrix} 1 & \gamma\psi & \psi \\ -\sigma & 1 & -\sigma\psi \\ -\gamma\sigma & \gamma & 1 \end{pmatrix} \\ \mathbf{Q} &= \mathbf{A}_0^{-1}\mathbf{A}_1 \\ &= \frac{1}{\gamma\sigma\psi + 1} \begin{pmatrix} 0 & \gamma\psi & \psi(\beta + \gamma\sigma) \\ 0 & \sigma(1 - \beta\psi) & 1 \\ 0 & \beta + \gamma\sigma & \gamma \end{pmatrix} \end{aligned}$$

and the two non-zero eigenvalues of \mathbf{Q} are

$$\begin{aligned} \kappa_1 &= \frac{1}{2(\gamma\sigma\psi + 1)} (1 + \beta + \gamma\sigma + \Psi), \\ \kappa_2 &= \frac{1}{2(\gamma\sigma\psi + 1)} (1 + \beta + \gamma\sigma - \Psi), \\ \Psi &= \sqrt{\beta^2 - 2\beta + \gamma^2\sigma^2 + 2\gamma\sigma + 2\gamma\sigma\beta - 4\gamma\sigma\beta\psi + 1}. \end{aligned} \tag{14}$$

Assuming that $|\kappa_i| < 1$ for $i = 1, 2$ then the solution is given by (9), which in this case is:

$$\begin{aligned} R_t &= \psi\pi_t + \varepsilon_{1t}, \\ x_t &= -\sigma R_t + \varepsilon_{2t}, \\ \pi_t &= \gamma x_t + \varepsilon_{3t}. \end{aligned} \tag{15}$$

This illustrates many of the features of DSGE models. First, the RE model parameter matrices, \mathbf{A}_0 and \mathbf{A}_1 , are written in terms of deeper parameters $\boldsymbol{\theta} = (\gamma, \sigma, \psi, \beta)'$. Second, the parameters which appear only in \mathbf{A}_1 do not enter the RE solution and, thus, do not enter the likelihood function. In this example, β does not appear in the likelihood function.⁶ Third, the restrictions necessary

⁶Notice, though, that ϕ which appears in \mathbf{A}_1 does appear in the likelihood function because it also appears in \mathbf{A}_0 .

to ensure regularity (i.e. $|\kappa_i| < 1$ for $i = 1, 2$), imply bounds involving the structural parameters, including the unidentified β . Thus, the parameter space is not variation free. Fourth, if β is fixed at some pre-selected value for the discount rate (as would be done by a calibrator), then the model is identified. Canova and Sala (2009) make similar points with a similar model.

2.3 DSGE models with lags

In order to reproduce the dynamics that are typically observed with macroeconomic data, most empirical DSGE models include lagged values of endogenous or exogenous (observed or unobserved) variables. For instance Clarida, Gali and Gertler (1999) assume that the errors in the IS and Phillips curve equations follow AR(1) processes and derive an optimal feedback policy for the interest rate based on the forecasts from these autoregressions. In this case, there is a predictable component in expected inflation because of the serial correlation in the equation errors.

Consider the special case of (1), where $\mathbf{A}_3 = \mathbf{\Phi}_u = 0$ so that the model only contains lagged endogenous variables

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{A}_1 E_t(\mathbf{y}_{t+1}) + \mathbf{A}_2 \mathbf{y}_{t-1} + \varepsilon_t. \quad (16)$$

In this case the unique solution is given by

$$\mathbf{y}_t = \mathbf{C} \mathbf{y}_{t-1} + \mathbf{A}_0^{-1} \varepsilon_t, \quad (17)$$

where \mathbf{C} solves the quadratic matrix equation $\mathbf{A}_1 \mathbf{C}^2 - \mathbf{A}_0 \mathbf{C} + \mathbf{A}_2 = \mathbf{0}$. The solution is unique and stationary if all the eigenvalues of \mathbf{C} and $(\mathbf{I} - \mathbf{A}_1 \mathbf{C})^{-1} \mathbf{A}_1$ lie strictly inside the unit circle. Therefore, the RE solution is observationally equivalent to the non-RE structural model :

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{A}_2 \mathbf{y}_{t-1} + \varepsilon_t,$$

where, in the case of the SEM, $\mathbf{C} = \mathbf{A}_0^{-1} \mathbf{A}_2$.

Again whereas the order condition for identification of the SEM requires n^2 restrictions, the RE model requires $2n^2$ restrictions. Not only is the RE model observationally equivalent to a purely backward looking SEM, it is observationally equivalent (in the sense of having the same reduced form), to any other model of expectations where in (16) $E_t(\mathbf{y}_{t+1})$ is replaced by $\mathbf{D} \mathbf{y}_{t-1}$. More specifically, knowing the form of the solution, (17), does not, on its own, provide information on the cross equation parametric restrictions. In either case, the identifying cross-equation restrictions are lost.

Thus, in models with lags, the same problem of observational equivalence between RE and other models recurs. One may be able to distinguish the reduced forms of particular RE models from other observationally equivalent models, because the RE models impose particular types of cross-equation restriction on the reduced form, which arise from the nature of the rational expectations. But such restrictions are subject to the objection made by Sims (1980), who

criticized identification by ‘incredible’ dynamic restrictions on the coefficients and lag lengths. RE models, which depend on restrictions on the form of the dynamics, such as AR(1) errors, are equally vulnerable to such objections.

2.3.1 Example 2: A Hybrid New Keynesian Phillips Curve (NKPC)

A specific example where it is well known how identification depends on assumptions about the dynamics is the hybrid NKPC with an exogenous driving process:

$$\pi_t = \beta_b \pi_{t-1} + \beta_f E_{t-1} \pi_{t+1} + \gamma x_t + \varepsilon_t, \quad (18)$$

where it is assumed that there is no feedback from π_t to x_t and that x_t can be written as a finite order autoregression.

The parameters of (18) are nonlinear functions of underlying structural parameters. For instance, following Gali, Gertler and Lopez-Salido (2005) suppose that there is staggered price setting, with a proportion of firms, $(1-\alpha)$, resetting prices in any period, and a proportion, α , keeping prices unchanged. Of those firms able to adjust prices only a fraction $(1-\omega)$ set prices optimally on the basis of expected marginal costs. A fraction ω use a rule of thumb based on lagged inflation. Then for a subjective discount factor, β , we have

$$\beta_f = \beta \alpha \delta^{-1}, \quad \beta_b = \omega \delta^{-1}, \quad \gamma = (1-\omega)(1-\alpha)(1-\lambda\alpha)\delta^{-1},$$

where $\delta = \alpha + \omega[1-\alpha(1-\beta)]$. If $\omega = 0$, all those who adjust prices do so optimally, then $\beta_f = \beta$, and $\beta_b = 0$. If the discount factor, $\beta = 1$, then $\beta_f + \beta_b = 1$ in either case. We will consider the identification of the three structural parameters,⁷ β_f , β_b , and γ ; but one could also consider identification of the four deeper parameters, β , α , ω , and δ . If the intermediate parameters are not identified, then the deeper parameters will not be.

If we assume ε_t is a martingale difference process; x_t follows a stationary time series process; there are no feedbacks from inflation to the output gap, $\beta_b, \beta_f \geq 0$, $\beta_f \beta_b \leq 1/4$ and $\beta_b + \beta_f \leq 1$, then the NKPC (18) has the unique solution,

$$\begin{aligned} \pi_t = & \kappa_b \pi_{t-1} + \frac{\gamma}{1 - \kappa_b \beta_f} \sum_{j=0}^{\infty} \kappa_f^{-j} E_{t-1}(y_{t+j}) \\ & + \frac{\gamma}{1 - \kappa_b \beta_f} [x_t - E_{t-1}(x_t)] + \frac{\varepsilon_t}{1 - \kappa_b \beta_f}, \end{aligned} \quad (19)$$

where κ_b and κ_f are roots of $\beta_f \kappa^2 - \kappa + \beta_b = 0$. The RE solution is unique if $|\kappa_b| \leq 1$ and $|\kappa_f| > 1$, which are satisfied if $\beta_b + \beta_f < 1$. In the case where $\beta_b + \beta_f = 1$, then $\kappa_b = 1$ and $\kappa_f = \beta_f^{-1}(1-\beta_f) > 1$ if $\beta_f < 1/2$. Inflation will be $I(1)$ in this case. Finally, if $\beta_b + \beta_f > 1$, the RE solution will be indeterminate and there exists a multiplicity of solutions. Analysis of identification in this latter case is beyond the scope of the present paper and will not be considered.

⁷These are sometimes called ‘semi-structural’, but this seems misleading since a parameter is either structural (i.e. invariant to policy) or not.

As noted originally in Pesaran (1981, 1987, Ch. 7) and emphasized recently by Mavroeidis (2005), Beyer et al (2007) and Nason and Smith (2008) among others, identification of the structural parameters critically depends on the process generating x_t . For example, suppose that x_t follows the $AR(1)$ process

$$x_t = \rho x_{t-1} + v_t. \quad (20)$$

Then the RE solution is given by

$$\pi_t = \alpha_1(\boldsymbol{\theta})\pi_{t-1} + \alpha_2(\boldsymbol{\theta})x_{t-1} + u_t, \quad (21)$$

where $\boldsymbol{\theta} = (\beta_b, \beta_f, \gamma, \rho)'$, $u_t = (\varepsilon_t + \gamma v_t) / (1 - \kappa_b \beta_f)$, and

$$\begin{aligned} \alpha_1 &= \alpha_1(\beta_b, \beta_f) = \kappa_b = \frac{1 - \sqrt{1 - 4\beta_f \beta_b}}{2\beta_f}, \\ \alpha_2 &= \alpha_2(\beta_b, \beta_f, \gamma, \rho) = \frac{1}{1 - \kappa_b \beta_f} \left(\frac{\gamma \rho}{1 - \rho \kappa_f^{-1}} \right) = \frac{\gamma \rho}{1 - \beta_f(\kappa_b + \rho)}, \end{aligned}$$

The reduced form for (π_t, x_t) is a restricted $VAR(1)$ that allows consistent estimation of the three parameters, α_1, α_2 , and ρ , whilst we have four unknown coefficients, β_f, β_b, γ , and ρ . In this case the structural parameters β_f, β_b and γ are not identified.

For identification we need the order of the $AR(p)$ process for the output gap to be at least equal to two. In general if the output gap, x_t , is $AR(p)$, the form for the RE solution is $ARDL(1, p)$ in π_t . In the case where x_t follows the $AR(2)$ process

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t,$$

then the extra instrument x_{t-2} exactly identifies the model. But the identification can be weak if ρ_2 is not sufficiently large. Weak instruments make GMM and the usual tests for over-identification unreliable, e.g., Stock, Wright and Yogo (2002). We return to this example below

3 Identification Issues in Rational Expectations DSGE Models: The Econometrics

3.1 Identification: General Issues

We begin with a brief overview of identification in a general context, before focussing on DSGE models. For the classical econometrician, a parameter is said to be identified if there is a well defined extremum of the objective function (such as likelihood function or GMM minimand) and not identified or under-identified if the objective function is flat. In the case where the objective function has little curvature, for instance because of weak instruments, there is said to be weak identification. There may be set identification, where the objective function is flat over a range, but one can put bounds on the parameter.

A model is globally identified if the well defined extremum of the objective function is a unique one in the entire parameter space. Local identification occurs if the extremum is unique locally. If the objective function is the likelihood function, then the information matrix is a measure of curvature. Let θ denote the parameters in a model. Rothenberg (1971) shows that local identification occurs at θ_0 if the information matrix is non-singular at θ_0 . In relation to weak identification, the information matrix can be used to measure the degree of curvature of the likelihood function (e.g. Iskrev, 2010b). The Jacobian of the mapping from reduced form to structural parameters can also be used to investigate local identification (e.g. Iskrev, 2010a).

In many cases one may have some parameters, α (e.g. reduced form parameters), which are functions of some deeper parameters, θ . Then it may be that α is identified, but elements of θ are not, e.g. if $\alpha = \theta_1\theta_2$. Typically, economists are interested in structural parameters, ones that are invariant to a class of policy interventions. As Marschak (1953) noted, for many economic purposes it may be easier, and equally useful for policy purposes, to identify policy invariant combinations of structural parameters, such as α , rather than the individual parameters themselves, a point Heckman (2010) also makes.

For the Bayesian econometrician, the objective function used to define identification is the likelihood function. Poirier (1998), building on earlier contributions (e.g. Kadane, 1974) sets out a framework for discussing identification in Bayesian models and we describe here a few of his key results which we will use later in our discussion of identification in DSGE models. We use notation where $\theta = (\theta_1, \theta_2)'$ is a vector of K parameters which lie in a region Φ , $p(\theta)$ is the prior, $p(\theta|y)$ is the posterior and $L(\theta; y)$ is the likelihood function. θ is identified if $L(\theta^{(1)}; y) = L(\theta^{(2)}; y)$ implies that $\theta^{(1)} = \theta^{(2)}$.

Consider the case where the parameters in θ_2 are identified but the scalar, θ_1 , is not.⁸ In DSGE models, the parameter space is often not variation free, so care must be taken with the bounds of the parameter space. Hence, we introduce notation where $\Lambda_1(\theta_2)$ and $\Lambda_2(\theta_1)$ defines the permissible range of values of θ_1 for given θ_2 , and θ_2 for given θ_1 , respectively. If the parameter space is variation free then we define $\Lambda_1 \equiv \Lambda_1(\theta_2)$ and $\Lambda_2 \equiv \Lambda_2(\theta_1)$.

Result 1: If θ_1 is not identified, then $L(\theta; y)$ is flat over $\theta_1 \in \Lambda_1(\theta_2)$ and the likelihood function can be written as depending only on θ_2 .

It is straightforward to use Result 1 and Bayes' theorem to show:

Result 2: If there is prior independence between θ_1 and θ_2 such that $p(\theta_1, \theta_2) = p(\theta_1)p(\theta_2)$ and the parameter space is a product space (i.e. $\Phi = \Lambda_1 \times \Lambda_2$) then $p(\theta_1|y) = p(\theta_1)$.

This is the commonly cited result that "posterior equals prior for unidentified parameters". Note, however, that this result only holds under prior independence and a variation free parameter space. If either of these conditions is not satisfied then $p(\theta_1|y) \neq p(\theta_1)$. Informally speaking, data based learning about θ_2 can "spill over" onto the unidentified θ_1 (see Koop and Poirier, 1997, for an example).

⁸Extensions to the case where θ_1 is a vector are straightforward.

As we shall discuss below, a better metric for investigating identification can be constructed based on Proposition 2 of Poirier (1998) which we state here.

Result 3: Let $p(\theta_1, \theta_2) = p(\theta_1|\theta_2)p(\theta_2)$ be the prior (which may exhibit correlation between θ_1 and θ_2), then

$$\begin{aligned} p(\theta_1|y) &= \int_{\Psi(\theta_2)} p(\theta_1|\theta_2, y) p(\theta_2|y) d\theta_2 \\ &= \int_{\Psi(\theta_2)} p(\theta_1|\theta_2) p(\theta_2|y) d\theta_2 \\ &= E_{\theta_2|y}[p(\theta_1|\theta_2)]. \end{aligned} \quad (22)$$

In words, the marginal posterior for the non-identified θ_1 will always be the posterior expectation of the conditional prior, $p(\theta_1|\theta_2)$.

The concepts discussed so far can be used with any econometric model, but we will use them below with DSGE models.

3.2 Bayesian Identification in DSGE Models

From the material in Section 2, it can be seen that some types of DSGE models are either simultaneous equations models (or closely related to them). For such models, of course, identification issues are well-understood. In the Bayesian literature on identification in the simultaneous equations model influential papers include Drèze (1976), Drèze and Richard (1983) and Kleibergen and van Dijk (1998). And Bayesian instrumental variable methods are well established (see, among many others, Kleibergen and Zivot, 2003, Hoogerheide, Kleibergen and van Dijk, 2007). Insofar as the DSGE model can be written as a conventional SEM, conventional methods can be used for Bayesian estimation and checking for identification. The NK-DSGE and NKPC models above fall in this category. There is also a literature relating to specific models such as the NKPC (e.g. Mavroeidis, 2005 and Kleibergen and Mavroeidis, 2009, 2010). For DSGEs which can be written in structural VAR form Rubio-Ramirez, Waggoner and Zha (2010) provides an exhaustive treatment. But, in general, with DSGE models direct verification of identification using such analytical methods is difficult.

In this paper, we will focus on empirical methods for determining identification (or weak identification). In the Bayesian context this means methods based on comparisons of priors and posteriors. For the classical econometrician, this often means methods based on the likelihood function.⁹ Iskrev (2010a) begins with the observation that normal likelihoods depend on the first two moments of the data. The Jacobian of the transformation from these first two moments to the structural parameters, θ , is crucial for identification.¹⁰ In particular, this Jacobian must be of full rank at θ_0 for the model to be locally identified at this point. Various choices for θ_0 can be made to investigate local identification at

⁹Canova and Sala (2009) consider matching impulse response functions, so the diagnostics for detecting the existence and source of identification problems are not expressed in likelihood terms. But the basic ideas and concepts they use transfer to likelihood function terms discussed below.

¹⁰The diagnostic recommended by Canova and Sala (2009) involves a quadratic form in the derivative of the VAR coefficients of the RE solution with respect to the structural parameters which is similar to this Jacobian.

different points in the parameter space (Iskrev takes a million draws from a prior to investigate identification over a wide region). Iskrev (2010a) recommends using analytical derivatives (since numerical differentiation can be inaccurate in these nonlinear transformations). This need for analytical derivatives adds an extra step in the coding process, but Iskrev (2010a) describes a relatively simple way of obtaining these derivatives. By checking whether the Jacobian matrix is of full rank, the researcher can see whether the model is identified. If the Jacobian is not of full rank an examination of where the rank deficiency occurs can shed light on the source of the identification failure. Iskrev (2010b) is similar in spirit to Iskrev (2010a), but focusses on strength of identification and uses the information matrix (which will reflect the curvature of the likelihood function) for this purpose. This approach also requires the calculation of analytical derivatives. More informally, directly looking at the likelihood function and whether it is flat (or nearly so) can reveal a lack of identification or weak identification and this is sometimes done. For instance, An and Schorfheide (2007, Figures 14 and 15) also present plots of the log-likelihood function.

It is worth noting that the methods in Canova and Sala (2009) and Iskrev (2010a,b) are not hypothesis testing procedures, but are better thought of as diagnostic procedures or indicators. Canova and Sala (2009) use the term diagnostics in reference to their methods. However, since diagnostic tests for misspecification are common in econometrics, we use the term indicator, to reinforce the point that these are not tests. Furthermore, these existing procedures can only check for local identification.

Bayesians typically use posterior simulation algorithms to estimate DSGE models. Our first proposed indicator can be calculated as part of such a posterior simulation algorithm without the need for additional steps such as the coding of analytical derivatives. Our second indicator involves using artificial data but it, as well, will involve standard posterior simulation algorithms. The Bayesian who uses proper priors will (under weak conditions) obtain a proper posterior, allowing for valid statistical inference. Since the parameters in DSGE models have a structural interpretation, sensible proper priors are usually available. These priors may be purely subjective or could reflect data from other sources (e.g. the priors could reflect estimates of structural parameters produced in microeconomic studies or could be based on a training sample of macroeconomic data). Given such prior information, there is a sense in which identification is not a worry for the Bayesian DSGE modeler.¹¹ However, if a parameter is not identified, then there is the possibility that there is no data-based learning about it and its posterior can solely reflect prior information. In complicated models such as DSGEs, where it can be hard to analytically disentangle identification issues, this can lead to the case where the researcher believes she is presenting posterior estimates but is really simply reproducing her prior.

Even if parameters are identified, weak identification can lead to relatively

¹¹In one of the classic Bayesian texts, identification is mentioned only in an oft-quoted footnote: “In passing it might be noted that unidentifiability causes no real difficulty in the Bayesian approach” (Lindley, 1971, page 46 footnote 34).

flat regions of the likelihood function where the prior is extremely influential. Such concerns have lead to a recent interest in identification issues in Bayesian DSGE modelling. Consider, for instance, Canova (2007, page 190) which states “while it is hard to ‘cheat’ in a classical framework, it is not very difficult to give the impression that identification problems are absent in a Bayesian framework by choosing tight enough priors, presenting well-behaved posterior distributions and entirely side-stepping the comparison between priors and posteriors”. In response to this, an increasingly common practice is to compare priors and posteriors for structural parameters, a practice which Canova (2007, page 191) refers to as “necessary [but] by no means sufficient” to reveal identification problems in DSGE models. We will draw on our earlier discussion of Bayesian identification (see Section 3.1) to discuss why this is so and introduce an alternative method for investigating identification in Bayesian DSGE models.

Result 2 of Section 3.1 underlies some informal discussion of identification in the Bayesian DSGE literature. For instance, An and Schorfheide (2007, page 127) say that: “A direct comparison of priors and posteriors can often provide valuable insights about the extent to which data provide information about parameters of interest.” This is true, but can be an imperfect way of formally investigating identification issues, since the posterior for an unidentified parameter can differ substantially from its prior if the non-identified parameter is, a priori, correlated with identified ones or if the parameter space is not a product space. Both of these are likely to hold with DSGE models. In fact, papers such as Del Negro and Schorfheide (2008) make a strong case that priors for structural parameters in DSGE models should not exhibit prior independence. Thus, informally comparing priors to posteriors could be useful to see *if* learning about parameters occurs, but may not be able tell the researcher *why* it is occurring. That is, the researcher may be unable to distinguish between learning via the likelihood function and learning solely due to the fact that the prior does not exhibit independence or the parameter space is not variation free.

Nevertheless, it is common in the Bayesian DSGE literature to use such informal comparisons of priors and posteriors, as the quote from An and Schorfheide above indicates. Among many others, Smets and Wouters (2007 page 594) compare prior and posteriors and note that the mean of the posterior distribution is typically quite close to the mean of the prior assumptions and later note that “It appears that the data are quite informative on the behavioral parameters, as indicated by the lower variance of the posterior distribution relative to the prior distribution.” As a recent example, Guerron-Quintana (2010, page 782) says “Initial estimation attempts showed that the posteriors of [certain structural parameters] sat on top of their priors. Hence those parameters are fixed to the values [taken from another paper]”. Statements similar to this implicitly suggest a comparison of prior to posterior is useful for checking identification in complicated DSGE models where it is not easy to analytically check identification.

Result 3 of Section 3.1 offers a promising way of formally investigating identification issues. In cases where a subset of structural parameters, say θ_2 , is known to be identified, but there is doubt regarding the identification of another

sub-set, θ_1 , then $p(\theta_1|y)$ should be equal to $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ where we use the notation of Section 3.1. For DSGE modelers interested in indicators that may shed light on identification issues, we would recommend comparing the properties of $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ in addition to (or instead of) comparing $p(\theta_1|y)$ to $p(\theta_1)$.

In terms of computation, note that our proposed indicator is typically easy to calculate. That is, the Bayesian DSGE modeler will typically be using an MCMC algorithm and, thus, posterior draws of θ_2 will be available. Calculation of $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ simply evolves evaluating $p(\theta_1|\theta_2)$ at each draw of θ_2 at a grid of values for θ_1 and averaging across the posterior draws of θ_2 . In many cases an analytical form for $p(\theta_1|\theta_2)$ will be available. For instance, if $p(\theta_1, \theta_2)$ is normal then $p(\theta_1|\theta_2)$ is also normal with textbook formula for its mean and variance. For priors which do not admit of analytical results, adding a prior simulation step at each posterior draw would only slightly add to the computational burden.

Formally, if a parameter is unidentified then $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ should be identical, apart from MCMC approximation error. Hence, the two densities cannot be used as a test for identification. That is, any difference between $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ beyond MCMC approximation error means identification is present. However, we can use the magnitude of the difference between these densities as an indicator revealing the strength of identification.

We will illustrate the usefulness of this indicator below. However, for DSGE models, it has one substantive drawback. For the theory underlying Result 3 to hold, the parameters in θ_2 must all be identified and θ_1 must not enter the likelihood function (this point is stressed on page 489 of Poirier, 1998). When working with an DSGE model, we would like to simply set θ to be the structural parameters. But typically we will not be able to do so (in the sense that the indicator defined in this way will not necessarily be zero for non-identified parameters). This is because the parameters in θ_2 defined in this way may not all be identified.

To see how this can happen in practice, we return to our example involving the NKPC (see Example 2 in Section 2) with an AR(1) process for the output gap. The two reduced form parameters in the inflation equations, a_1 and a_2 , depend on three structural parameters β_f, β_b and γ and, thus, there is an identification problem involving these three structural parameters.¹² In contrast to Example 1, the identification problem does not manifest itself simply in terms of a single structural parameter which does not enter the likelihood function. Using the notation of Result 3, we might be tempted to set $\theta_1 = \beta_f$ and $\theta_2 = (\beta_b, \gamma, \rho)$ in order to investigate the identification of β_f . However, it can be easily seen that the derivations in (22) used to prove Result 3 are no longer valid. In such cases, we will not have $p(\theta_1|y) = E_{\theta_2|y}[p(\theta_1|\theta_2)]$, even though θ_1 is only partially identified.

The advice given by Poirier (1998) in such cases is to re-parameterize the model so that θ_2 contains only identified parameters. In some DSGE cases, a

¹² a_2 also depends on ρ but, since this parameter is identified through the AR(1) process for the output gap, this dependence is irrelevant for the argument we are making here.

simple way of choosing θ_2 suggests itself: let θ_2 be some or all of the reduced form parameters and θ_1 be one of the structural parameters. We know the reduced form parameters are identified and, thus, the conditions under which Result 3 holds are satisfied. In such cases, we can recommend a comparison of $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ as shedding light on the identification of θ_1 . The drawback of this strategy is that you have to know about the identification of the model in advance. Thus it will be most useful as an indicator of the strength of identification of each parameter rather than an indicator of whether identification occurs or not. We will return to these points in our empirical illustrations.

3.3 A Bayesian Identification Indicator Based on Large Sample Results

The advantage of the identification indicator of the preceding sub-section is that it can be calculated as a by-product of estimating a DSGE model (on the actual data) using MCMC methods. The disadvantage is that we typically cannot simply work with the structural parameters of a model. A second Bayesian identification indicator can be obtained based on the asymptotic theory of non-identified models written in terms of the structural parameters. Thus, we can focus on identification of a single structural parameter without worrying about whether the other structural parameters are identified or not. Empirically, this indicator involves simulating artificial data sets of increasing size and then estimating the DSGE model using these data sets. Since the generation of simulated data is fairly standard in the DSGE literature, this strategy fits in with existing empirical methodologies.

To explain the theory underlying this second indicator of identification, note that standard Bayesian results for stationary models (see, e.g., Berger, 1985, page 224), imply that, under certain regularity conditions, the Bayesian asymptotic theory relating to the posterior is numerically identical to the asymptotic distribution theory for the maximum likelihood estimator. Thus, for instance, the posterior for θ will asymptotically converge to its true value and the role of the prior will vanish. One of the regularity conditions is that θ is identified. In this sub-section, we relax this assumption and show that this asymptotic convergence will not occur. This result holds even for cases of partial identification such as the NKPC of our Example 2.

To derive this result, let $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ be a $p \times 1$ vector of structural parameters of interest in a DSGE model of the type set out in Section 2. Suppose that the likelihood function for a sample of T observations can be written as $L_T(\alpha(\theta); y)$ where $\alpha(\cdot)$ is a $k \times 1$ vector-valued function of θ , with at least first and second order derivatives, $\frac{\partial \alpha(\theta)}{\partial \theta_i}$, and $\frac{\partial^2 \alpha(\theta)}{\partial \theta_i \partial \theta_j}$. Denote the $k \times p$ matrix of derivatives $\frac{\partial \alpha(\theta)}{\partial \theta'}$ by $R(\theta)$. Let $\ell_T(\alpha) = \ln [L_T(\alpha(\theta); y)]$, and assume that

$$Q_T(\alpha) = \frac{-1}{T} \frac{\partial^2 \ell_T(\alpha)}{\partial \alpha \partial \alpha'}$$

is a positive definite matrix for all values of $\alpha \in \mathcal{A}$, and $\text{plim}_{T \rightarrow \infty} Q_T(\alpha) = Q > 0$. Denote the maximum likelihood estimator of α by $\hat{\alpha}_T$ and assume that

$$\sqrt{T}(\hat{\alpha}_T - \alpha_0) \rightarrow_d N(0, Q^{-1}),$$

where α_0 is the true value of $\alpha \in \mathcal{A}$.

The object of the exercise is to derive the posterior distribution of θ_1 assuming that the prior density of θ is given by

$$p(\theta) = (2\pi)^{-p/2} |\underline{\mathbf{H}}|^{1/2} \exp -\frac{1}{2}(\theta - \underline{\theta})' \underline{\mathbf{H}}(\theta - \underline{\theta}), \quad (23)$$

where $\underline{\theta}$ and $\underline{\mathbf{H}}$ are prior mean vector and prior precision matrix of θ . The posterior density of θ is given by

$$p(\theta | y) = \frac{p(\theta) L_T(\alpha(\theta); y)}{p(y)},$$

or

$$p(\theta | y) = \exp\{\ln[p(\theta)] + \ell_T(\alpha(\theta)) - \ln(p(y))\}. \quad (24)$$

When T is finite, assuming a proper prior, the posterior of θ is well defined even if θ is not identified. In the classical sense θ is globally identified if $\text{Rank}(R(\theta)) = k \geq p$ for all $\theta \in \Theta$, and θ is identified locally in the neighborhood of $\underline{\theta}$ if $\text{Rank}(R(\underline{\theta})) = k \geq p$. A necessary condition for identification is given by $p \leq k$.

Suppose that $p = k + 1$, and θ_1 is non-identified. But for simplicity assume that the remaining $(p - 1) \times 1$ structural parameters, $\theta_* = (\theta_2, \dots, \theta_p)'$, are identified. The generalization to the case where two or more structural parameters are unidentified is discussed below.

Note that, if θ_1 does not enter the likelihood function and θ_1 is, a priori, uncorrelated with θ_* , then Result 2 of Section 3.1 says that $p(\theta_1) = p(\theta_1 | y)$ for all T . In this case, the posterior precision of θ_1 does not get updated at all as sample size increases (i.e. it remains equal to prior precision). Formally, if we define average posterior precision as posterior precision divided by T , then average posterior precision will tend to zero with T in this case. The derivations below show that the statement “average posterior precision will tend to zero with T ” will hold whenever θ_1 is unidentified (even if the nature of the identification problem is more complicated than that assumed in Result 2). However, when θ_1 is identified, then the average posterior precision will tend to a strictly positive constant which is independent of the prior precision.

To see why the result arises, note that although θ is not identified, there exists $\hat{\theta}_T$ such that $\hat{\alpha}_T = \alpha(\hat{\theta}_T)$. The choice of $\hat{\theta}_T$ is not unique but as we shall see this is of no consequence for the derivation of posterior precision of θ_1 .

Consider now the following Taylor series expansion of $\ell_T(\alpha(\theta))$ around $\hat{\theta}_T$.

$$\begin{aligned}\ell_T(\alpha(\theta)) &= \ell_T(\alpha(\hat{\theta}_T)) + \sum_{i=1}^p \sum_{s=1}^k \frac{\partial \ell_T(\hat{\alpha}_T)}{\partial \alpha_s} \frac{\partial \alpha_s(\hat{\theta}_T)}{\partial \theta_i} (\theta_i - \hat{\theta}_{iT}) + \\ &\quad \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \sum_{s=1}^k \sum_{r=1}^k \frac{\partial^2 \ell_T(\hat{\alpha}_T)}{\partial \alpha_r \partial \alpha_s} \frac{\partial \alpha_r(\hat{\theta}_T)}{\partial \theta_i} \frac{\partial \alpha_s(\hat{\theta}_T)}{\partial \theta_j} (\theta_i - \hat{\theta}_{iT})(\theta_j - \hat{\theta}_{jT}) \\ &\quad + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \sum_{s=1}^k \frac{\partial \ell_T(\hat{\alpha}_T)}{\partial \alpha_s} \frac{\partial^2 \alpha_s(\hat{\theta}_T)}{\partial \theta_i \partial \theta_j} (\theta_i - \hat{\theta}_{iT})(\theta_j - \hat{\theta}_{jT}) + \text{higher order terms}.\end{aligned}$$

Assuming that the higher order terms are negligible and noting that $\frac{\partial \ell_T(\hat{\alpha}_T)}{\partial \alpha_s} = 0$, for all s , we have

$$\ell_T(\alpha(\theta)) \approx \ell_T(\alpha(\hat{\theta}_T)) - \frac{T}{2} (\theta - \hat{\theta}_T)' \hat{S}_T (\theta - \hat{\theta}_T)$$

where $\hat{S}_T = \hat{R}' Q_T \hat{R}$ and $\hat{R} = R(\hat{\theta}_T)$. Using this result and (23) in (24) we have

$$p(\theta | y) \propto \exp\left\{-\frac{1}{2}(\theta - \underline{\theta})' \underline{H}(\theta - \underline{\theta}) - \frac{T}{2}(\theta - \hat{\theta}_T)' \hat{S}_T (\theta - \hat{\theta}_T)\right\}.$$

Using textbook results for combining a normal prior with normal likelihood, the posterior distribution of θ is approximately normal with mean $\bar{\theta}_T$ and the precision matrix \bar{H}_T , where

$$\bar{\theta}_T = \left(T \hat{S}_T + \underline{H}\right)^{-1} (T \hat{S}_T \hat{\theta}_T + \underline{H} \underline{\theta}), \text{ and } \bar{H}_T = T \hat{S}_T + \underline{H}.$$

It is clear that $\bar{\theta}_T$ is defined even if \hat{R} , or equivalently \hat{S}_T , fails the rank condition. Since the marginals of multivariate normal are also normally distributed, the posterior of θ_1 is (approximately) normally distributed with mean $\bar{\theta}_{1T}$, where $\bar{\theta}_{1T}$ is the first element of $\bar{\theta}_T$, and the posterior precision of θ_1 is (approximately) given by (suppressing the T subscript to simplify the exposition)

$$\bar{h}_{11} = \bar{H}_{11} - \bar{H}_{12} \bar{H}_{22}^{-1} \bar{H}_{21},$$

where

$$\bar{H} = \begin{pmatrix} \bar{H}_{11} & \bar{H}_{12} \\ \bar{H}_{21} & \bar{H}_{22} \end{pmatrix} = T \begin{pmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & \hat{S}_{22} \end{pmatrix} + \begin{pmatrix} \underline{H}_{11} & \underline{H}_{12} \\ \underline{H}_{21} & \underline{H}_{22} \end{pmatrix}.$$

Hence

$$\bar{h}_{11} = (T \hat{S}_{11} + \underline{H}_{11}) - \left(T \hat{S}_{12} + \underline{H}_{12}\right) \left(T \hat{S}_{22} + \underline{H}_{22}\right)^{-1} \left(T \hat{S}_{21} + \underline{H}_{21}\right).$$

It is clear that when T is finite \bar{h}_{11} is well defined irrespective of whether \hat{R} is a full rank matrix or not. Note that, even if $\underline{H}_{21} = 0$, the posterior of θ_1 may

not be independent of the posterior of θ_* . This is because in general $\hat{S}_{12} \neq 0$. In the case where $p = k$ posterior independence follows if $R_{12} = Q_{12} = 0$.

Consider now the case where $T \rightarrow \infty$, and note that since

$$T^{-1}\bar{h}_{11} = (\hat{S}_{11} + T^{-1}\mathbb{H}_{11}) - (\hat{S}_{12} + T^{-1}\mathbb{H}_{12}) (\hat{S}_{22} + T^{-1}\mathbb{H}_{22})^{-1} (\hat{S}_{21} + T^{-1}\mathbb{H}_{21}),$$

then as $T \rightarrow \infty$ we have

$$\lim_{T \rightarrow \infty} (T^{-1}\bar{h}_{11}) = S_{11} - S_{12}S_{22}^{-1}S_{21},$$

where

$$S = p \lim (\hat{R}'Q_T\hat{R}) = R'QR = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

In the case where R is full rank, S is a positive definite matrix. Hence, $S_{11} - S_{12}S_{22}^{-1}S_{21} > 0$, and $\lim_{T \rightarrow \infty} (T^{-1}\bar{h}_{11})$ is strictly positive and does not depend on the prior precision.

But when the rank condition is not satisfied the above result does not follow. For example, suppose that $p = k + 1$ and consider the following partition of R

$$R = \begin{pmatrix} r & R_* \end{pmatrix},$$

where r is a $k \times 1$ vector and R_* is $k \times k$. Recall that $p = 1 + k$, and by assumption R_* is a non-singular matrix. Then

$$S = R'QR = \begin{pmatrix} r' \\ R_*' \end{pmatrix} Q \begin{pmatrix} r & R_* \end{pmatrix} = \begin{pmatrix} r'Qr & r'QR_* \\ R_*'Qr & R_*'QR_* \end{pmatrix}.$$

Hence

$$S_{11} - S_{12}S_{22}^{-1}S_{21} = r'Qr - r'QR_*(R_*'QR_*)^{-1}R_*'Qr.$$

But since R_* and Q are both non-singular then it readily follows that $S_{11} - S_{12}S_{22}^{-1}S_{21} = 0$, and

$$\lim_{T \rightarrow \infty} (T^{-1}\bar{h}_{11}) = 0.$$

Namely the posterior precision of θ_1 must change at a rate slower than T when θ_1 is non-identified. In terms of the posterior variance this result confirms that the posterior variance of a non-identified parameter need not tend to zero, and in cases that it does its rate of decline must be slower than T .

The above result readily generalizes when two or more of the structural parameters are unidentified. Consider the case where $\theta = (\theta'_1, \theta'_2)'$ with θ_1 the $s \times 1$ vector of unidentified parameters and θ_2 the $(p - s) \times 1$ vector of the identified parameters where $p - s = k$. Partition R as $R = (R_1, R_2)$, where R_1 and R_2 are $k \times s$ and $k \times k$ matrices, where R_2 is non-singular and write S as

$$S = R'QR = \begin{pmatrix} R_1'QR_1 & R_1'QR_2 \\ R_2'QR_1 & R_2'QR_2 \end{pmatrix}.$$

Then noting that by assumption Q and R_2 are non-singular matrices, it readily follows that $S_{11} - S_{12}S_{22}^{-1}S_{21} = 0$, as required.

We propose to use these results as a second Bayesian indicator of identification which can be used when our first Bayesian indicator of identification is not applicable. That is, we recommend the following strategy: Suppose that it is of interest to investigate if one or more elements of θ are identified, for example in the locality of the prior mean vector, $\underline{\theta}$. First, the researcher should generate an artificial data set of size T from the DSGE model at $\theta = \underline{\theta}$. T should be chosen to be a large value where asymptotic results are expected to be very good approximations (e.g. the empirical illustrations below set $T = 10,000$). Second, the researcher should estimate the DSGE model using sample sizes $\tau = cT$ for a grid of values for c (e.g. $c = 0.0001, 0.001, 0.01, 0.1, 1.0$) and calculate a measure which relates to the posterior precision (e.g. the posterior variance) for every parameter. By comparing the behavior of the measure over different sample sizes, the researcher can see which parameters are identified and which are not. For instance, the posterior variance should be going to zero with sample size for identified parameters, but will not be doing so for unidentified parameters.

Note that this strategy will be an indicator of local identification (i.e. it will check identification at the parameter values used to generate the artificial data). Hence, the researcher may wish to carry out the procedure for various artificial data sets generated with different parameter values. This strategy is comparable to the one used by Iskrev (2010a), who draws parameter values from the prior and checks identification at each of the draws.

3.3.1 Example: Regressions with exactly collinear regressors

As an example of the preceding derivations, consider the following simple regression model

$$y_t = \theta_1 x_{1t} + \theta_2 x_{2t} + u_t,$$

where θ_1 and θ_2 are the parameters of interest. Suppose that $x_{2t} = \rho x_{1t}$ where ρ is a known non-zero constant. Then

$$y_t = \alpha(\theta) x_{1t} + u_t,$$

where $\alpha = \alpha(\theta) = \theta_1 + \rho\theta_2$. Assuming a normal prior for the parameters of interest, as in (23), causes α and θ_1 to be dependent on one another (unless θ_1 and θ_2 are a priori dependent and $\rho = -v_{11}/v_{12}$, where v_{ij} for $i, j = 1, 2$ denote the prior variance-covariances of θ_1 and θ_2). Therefore, in cases where θ_1 and θ_2 are a priori independent (i.e. $v_{12} = 0$) or when $v_{12} \neq 0$ but $\rho \neq -v_{11}/v_{12}$, then α and θ_1 are a priori dependent and Result 2 of Section 3.1. does not apply.

Using our previous results, the posterior precision of θ_1 for a finite T is given by

$$T^{-1}\bar{h}_{11} = (\hat{s}_{11} + T^{-1}\bar{h}_{11}) - (\hat{s}_{12} + T^{-1}\bar{h}_{12}) (\hat{s}_{22} + T^{-1}\bar{h}_{22})^{-1} (\hat{s}_{21} + T^{-1}\bar{h}_{21}),$$

where

$$\hat{S} = R'Q_T R, R = \begin{pmatrix} 1 & \rho \end{pmatrix}, Q_T = T^{-1} \sum_{t=1}^T x_{1t}^2 = s_{T1}^2 = s_1^2 > 0,$$

and hence

$$\hat{s}_{11} = s_1^2, \hat{s}_{12} = \hat{s}_{21} = \rho s_1^2 \text{ and } \hat{s}_{22} = \rho^2 s_1^2.$$

Therefore

$$\begin{aligned} T^{-1}\bar{h}_{11} &= s_1^2 \left\{ 1 + T^{-1}(\mathbf{h}_{11}/s_1^2) - \frac{(\rho + T^{-1}\mathbf{h}_{21}/s_1^2)^2}{\rho^2 + T^{-1}(\mathbf{h}_{22}/s_1^2)} \right\} \\ &= T^{-1}s_1^2 \left\{ \frac{(\mathbf{h}_{22}/s_1^2) + (\mathbf{h}_{11}/s_1^2)\rho^2 + (\mathbf{h}_{11}/s_1^2)T^{-1}(\mathbf{h}_{22}/s_1^2) - 2\rho\mathbf{h}_{21}/s_1^2 - T^{-1}(\mathbf{h}_{21}/s_1^2)^2}{\rho^2 + T^{-1}(\mathbf{h}_{22}/s_1^2)} \right\}. \end{aligned}$$

When the priors of θ_1 and θ_2 are independent the above expression simplifies to

$$\bar{h}_{11} = \frac{\mathbf{h}_{22} + \rho^2\mathbf{h}_{11} + T^{-1}\mathbf{h}_{11}\mathbf{h}_{22}/s_1^2}{\rho^2 + T^{-1}(\mathbf{h}_{22}/s_1^2)}.$$

Hence, posterior precision (\bar{h}_{11}) of the unidentified parameter, θ_1 , differs from its prior precision (\mathbf{h}_{11}) for all T , and as $T \rightarrow \infty$, even though θ_1 and θ_2 are assumed to be a priori independent. For T sufficiently large we have

$$\lim_{T \rightarrow \infty} \bar{h}_{11} = \mathbf{h}_{11} + \rho^{-2}\mathbf{h}_{22}.$$

Hence, the posterior precision is bounded in T , in contrast to the posterior precision of an identified parameter.

The extent to which the posterior precision deviates from the prior precision is determined by \mathbf{h}_{22}/ρ^2 . It is also worth noting, however, that as T increases the posterior precision declines. This could be viewed as an indication that θ_1 is not identified. In the case where a parameter is identified we would expect the posterior precision to rise with T and eventually dominate the prior precision.

3.3.2 Example: The NKPC with no Backward Looking Behavior

Consider the NKPC (see Example 2 in Section 2.3.1) and, for simplicity, assume that there is no backward looking behavior ($\beta_b = 0$) and simplify notation by defining $\beta \equiv \beta_f$. The solution of this model can be written as:

$$y_t = \alpha(\gamma, \beta)x_t + u_t,$$

where

$$\alpha \equiv \alpha(\gamma, \beta) = \frac{\gamma\rho}{1 - \rho\beta},$$

and assume that ρ is known (it is identified, so asymptotically its estimator will converge to the true value). For this example we have

$$R = \left(\frac{\rho}{1 - \rho\beta} \quad \frac{\gamma\rho^2}{(1 - \rho\beta)^2} \right).$$

It is clear that the rank condition is not satisfied and neither of the structural parameters is identified.

This example is complicated by the fact that R depends on the unknown parameters. With this in mind we note that R can also be written as

$$R = \frac{\rho}{1 - \rho\hat{\beta}} \begin{pmatrix} 1 & \alpha \end{pmatrix}.$$

Furthermore

$$\begin{aligned} \hat{R} &= \frac{\rho}{1 - \rho\hat{\beta}} \begin{pmatrix} 1 & \hat{\alpha} \end{pmatrix}, \quad Q_T = T^{-1} \sum_{t=1}^T x_t^2 = s_x^2 > 0, \\ \hat{S} &= \hat{R}' Q_T \hat{R} = \frac{\rho^2 s_x^2}{(1 - \rho\hat{\beta})^2} \begin{pmatrix} 1 & \hat{\alpha} \\ \hat{\alpha} & \hat{\alpha}^2 \end{pmatrix}. \end{aligned}$$

$\hat{\alpha}$ could be selected to be the OLS estimator of α in the regression of y_t on x_t . But the choice of $\hat{\beta}$ is arbitrary so long as it lies in the range of $[0, 1]$ - but as we shall see below in the limit $\hat{\beta}$ gets eliminated from the posterior precisions.

It can now be seen that the following approximate results hold:

$$T^{-1} \bar{h}_{\gamma\gamma} \approx \left(\frac{\rho^2 s_x^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\gamma\gamma} \right) - \frac{\left(\frac{\rho^2 s_x^2 \hat{\alpha}}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\gamma\beta} \right)^2}{\frac{\rho^2 s_x^2 \hat{\alpha}^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\beta\beta}},$$

and similarly

$$T^{-1} \bar{h}_{\beta\beta} \approx \frac{\rho^2 \hat{\alpha}^2 s_x^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\beta\beta} - \frac{\left(\frac{\rho^2 s_x^2 \hat{\alpha}}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\gamma\beta} \right)^2}{\frac{\rho^2 s_x^2 \hat{\alpha}^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\gamma\gamma}}.$$

In the case where priors of γ and β are independent of one another, $\mathbf{h}_{\gamma\beta} = 0$, we obtain

$$\bar{h}_{\gamma\gamma} \approx \mathbf{h}_{\gamma\gamma} + \frac{\frac{\rho^2 s_x^2}{(1 - \rho\hat{\beta})^2} \mathbf{h}_{\beta\beta}}{\frac{\rho^2 s_x^2 \hat{\alpha}^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\beta\beta}}$$

and as $T \rightarrow \infty$

$$\lim_{T \rightarrow \infty} (\bar{h}_{\gamma\gamma}) = \mathbf{h}_{\gamma\gamma} + \frac{\mathbf{h}_{\beta\beta}}{\alpha^2},$$

which does not depend on the nuisance parameter β . Similarly

$$\bar{h}_{\beta\beta} \approx \mathbf{h}_{\beta\beta} + \frac{\frac{\rho^2 \hat{\alpha}^2 s_x^2}{(1 - \rho\hat{\beta})^2} \mathbf{h}_{\gamma\gamma}}{\frac{\rho^2 s_x^2}{(1 - \rho\hat{\beta})^2} + T^{-1} \mathbf{h}_{\gamma\gamma}},$$

and

$$\lim_{T \rightarrow \infty} (\bar{h}_{\beta\beta}) = \mathbf{h}_{\beta\beta} + \alpha^2 \mathbf{h}_{\gamma\gamma}.$$

In the case of both structural parameters the posterior precisions are dominated by the prior precisions even if T is sufficiently large. In neither case do the prior and the posterior precisions coincide despite the prior independence of the structural parameters.

This example illustrates the drawback of our first Bayesian indicator of identification discussed above. That is, Results 2 and 3 of Section 3.1 (or the propositions of Poirier, 1998, Section 2), obtain when the vector of parameters is written in terms of those which are identified and those which are not. In cases of partial identification such as this one, the vector of structural parameters cannot be written in this way (i.e. β and γ are both unidentified and thus, neither can be included in what we called θ_2 in Results 2 and 3). However, it also shows how our large sample derivations can be used as a second Bayesian indicator of identification which is valid even in cases such as this.

4 Applications

In this section, we illustrate both of our Bayesian identification indicators in the context of the two examples of DSGE models introduced in Section 2. These are the NK-DSGE (see Section 2.2.1) and the NKPC (see Section 2.3.1). For the NKPC, we use the simplified version of the model with no backward looking behavior (used in Section 3.3.2).

4.1 Example 1: Bayesian identification of the simple NK-DSGE model

Previously, we introduced a simple NK-DSGE in (11), (12) and (13). We will illustrate some issues relating to Bayesian inference and identification in this simple and easily understood model where the identification of the model can be immediately seen. This example involves four structural parameters, σ , γ , ψ and β . The rational expectations solution given in (15) does not involve β so this parameter is unidentified. However, the bounds given in (14) which ensure regularity such that there is a unique stationary solution do involve β .

We generated one artificial data set of $T = 10,000$ observations from (15) with $\sigma = 0.4$, $\gamma = 0.75$ and $\psi = 2.0$. These values were chosen so as to be not too far from the boundaries given in (14), but also not too near.¹³ The errors, ε_{jt} for $j = 1, 2, 3$ are all standard normal and independent of one another.

We estimate the model using different sample sizes and two different priors. Both priors are normal with prior means: $E(\sigma) = 0.4$, $E(\gamma) = 0.75$, $E(\psi) = 2.0$ and $E(\beta) = 0.9$. The two priors differ in their prior variances. Let $\theta = (\sigma, \gamma, \psi, \beta)'$. The first prior (which we call the Independent Prior) has $\text{var}(\theta) = I$. The second prior (the Dependent Prior) has the same prior covariance matrix except for a single element: this is the covariance between β and σ which is set to 0.9. These priors are combined with the likelihood function based on the three

¹³The posterior simulation algorithm rejected 3.1% of the draws for violating the bounds.

equation system in (15). We use a random walk Metropolis-Hastings algorithm to do posterior simulation in this model.¹⁴

We begin by illustrating the properties of our first Bayesian identification indicator with $T = 100$ (i.e. we use only the first 100 of the artificially generated observations). Figures 1 and 2 graph various priors and posteriors for β and σ , respectively, for the Independent Prior. Figures 3 and 4 do the same for the Dependent Prior. For the sake of brevity, we only present graphs for one identified and one non-identified parameter and, thus, do not present results for γ and ψ .

Consider first the priors and posteriors for β . Since β is unidentified, a naive researcher may expect its posterior to equal its prior. For the reasons discussed in Section 3, this may not be the case. The top panels of Figures 1 and 3 illustrate this empirically. Even with the Independent Prior (where β is, a priori, uncorrelated with the other parameters in the model), the fact that β enters the bounds for the regularity region given in (14) has an appreciable impact on the posterior in Figure 1. In Figure 3 (which uses a prior where the unidentified β is strongly correlated with the identified σ), this effect is even more noticeable. The posterior for β has a much smaller variance than its prior, indicating how information about σ is spilling over onto β .

The priors and posterior for σ show (as expected) that learning is occurring about this identified parameter. The posteriors in the top panels of Figures 2 and 4 are concentrated near the true value used to generate the data.

This example illustrates an important point we have made previously: An informal comparison of priors and posteriors of structural parameters in DSGE models can be a useful way of investigating if learning is occurring about a parameter. However, such a comparison will not tell the researcher why the learning is occurring. Our figures show posteriors can differ from priors, even for parameters which do not enter the likelihood function, either when the parameter space is not variation free or through prior correlations with identified parameters. Since DSGE models will often exhibit such features, this illustration shows how caution should be taken when interpreting comparisons of priors with posteriors.

In Section 3.2, we recommended using an alternative indicator based on (22). If interest centers on identification issues relating to θ_1 then this indicator involved comparing $p(\theta_1|y)$ to $E_{\theta_2|y}[p(\theta_1|\theta_2)]$. The bottom panels of Figures 1 through 4 present such a comparison for β and σ for our two priors. Clearly our indicator is working well. For the non-identified parameter, $p(\beta|y)$ and $E_{\sigma,\gamma,\psi|y}[p(\beta|\sigma,\gamma,\psi)]$ are the same density. For the identified parameter, $p(\sigma|y)$ and $E_{\beta,\gamma,\psi|y}[p(\sigma|\beta,\gamma,\psi)]$ are massively different, indicating the parameter is strongly identified.

¹⁴In this small model, with only four parameters, this algorithm works well. In larger models, depending on the form of the prior, more efficient posterior simulation algorithms could be used.

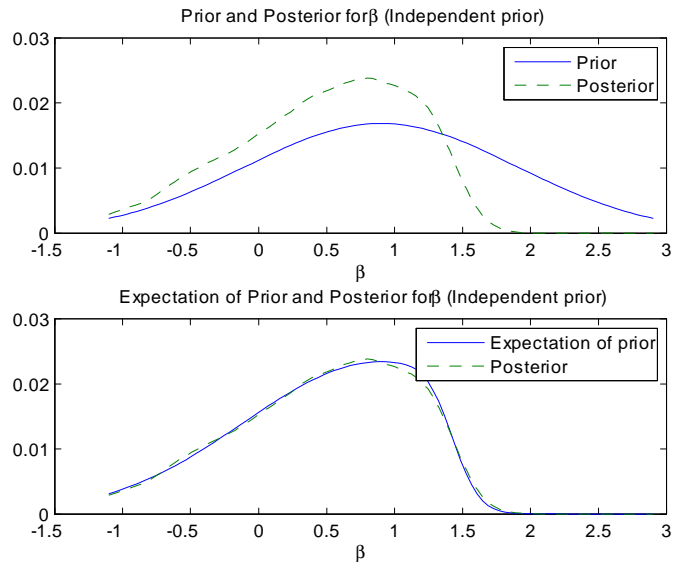


Figure 1

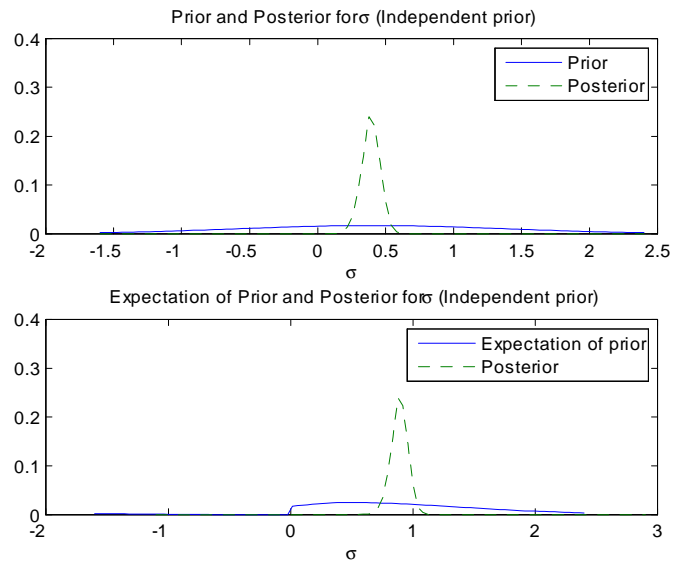


Figure 2

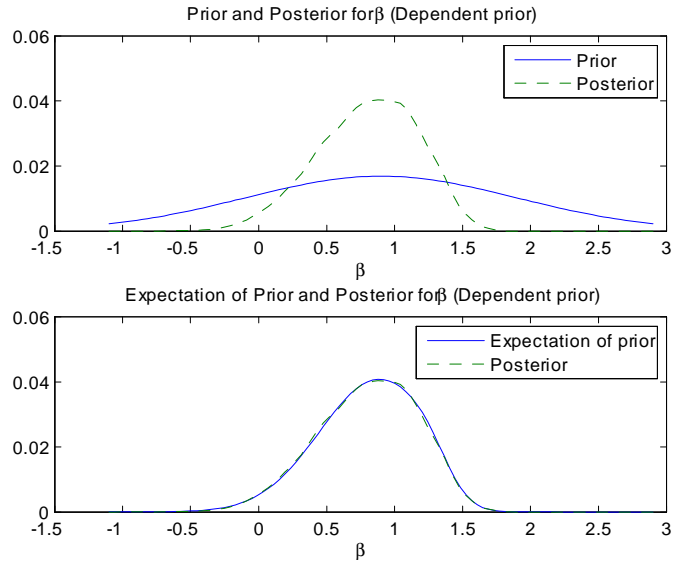


Figure 3

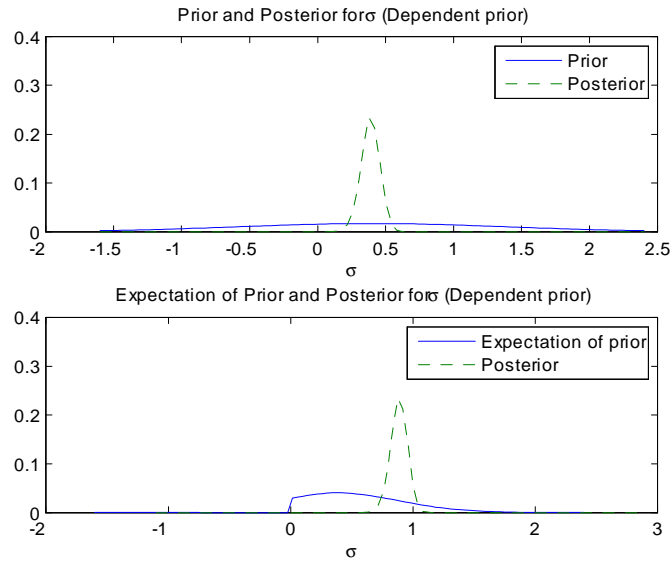


Figure 4

To illustrate our second Bayesian identification indicator, based on large sample derivations, Table 1 presents the posterior variances of the parameters

in the model using larger and larger data sets. Remember that the theoretical derivations underlying our second Bayesian identification indicator imply that the posterior variance of the identified parameters, σ , γ and ψ , should be tending to zero at the rate of $1/T$. But the posterior variance of the unidentified parameter, β , will go to zero (if at all) at a slower rate. These properties can be clearly seen in Table 1. In contrast to the identified parameters, the posterior variance of β is still substantial even for a sample of size 10,000. This result holds irrespective of whether the prior distribution of β depends on the other parameters or not. For example, in the case of independent priors, when $T = 10,000$ the posterior variance of β is 0.411 as compared to 3×10^{-5} , 1×10^{-4} , and 1×10^{-4} for σ , γ and ψ , respectively.

Table 1: Posterior Variances of Structural Parameters				
Number of observations	σ	γ	ψ	β
Independent Prior				
T=10	0.101	0.479	0.156	0.516
T=20	0.042	0.215	0.080	0.467
T=50	0.009	0.027	0.035	0.457
T=100	0.005	0.013	0.017	0.432
T=1,000	4×10^{-4}	0.001	0.002	0.426
T=10,000	3×10^{-5}	1×10^{-4}	1×10^{-4}	0.411
Dependent Prior				
T=10	0.071	0.321	0.162	0.298
T=20	0.043	0.222	0.080	0.190
T=50	0.010	0.027	0.036	0.153
T=100	0.005	0.014	0.017	0.139
T=1,000	3×10^{-4}	0.001	0.002	0.132
T=10,000	4×10^{-5}	1×10^{-4}	2×10^{-4}	0.128

4.2 Example 2: The NKPC model

For the reasons discussed in Section 3.2 and in the theoretical derivations of Section 3.3.2, our first Bayesian indicator of identification will not work reliably when we work with the NKPC and parameterize the model in terms of its structural parameters. However, our second Bayesian identification indicator of Section 3.3, based on large sample theory, should still work. Accordingly, we use the NKPC (as in Section 3.3.2, we assume there is no backward looking behavior) to investigate the performance of this second identification indicator. We consider both an unidentified and identified version of the NKPC. The unidentified version assumes an AR(1) process for the output gap. The identified version assumes an AR(2) process for the output gap. Including the identified version allows us to investigate issues relating to the strength of identification.

4.2.1 The NKPC with AR(1) Process for the Output Gap

When no backward looking behavior exists ($\beta_b = 0$), then the hybrid NKPC with AR(1) process, (20) is parameterized in terms of three structural parameters β, γ and ρ (where β is the forward-looking coefficient in the NKPC). The RE solution given in (21) simplifies and depends on two reduced form parameters, α and ρ . The lack of identification reveals itself through the mapping from structural to reduced form parameters: $\alpha = \rho\gamma/(1 - \rho\beta)$.

We generated one artificial data set of 10,000 observations from the NKPC with $\beta = 0.6, \gamma = 0.9, \rho = 0.3$. In addition, u_t is i.i.d. $N(0, 0.25)$ and v_t is i.i.d. $N(0, 1)$. For γ and ρ , the prior is normal: $N(0.5, 0.1\mathbf{I}_2)$. This prior is chosen so that the prior means are a bit different (but not too different) from the true values and prior variance is fairly informative. For β , we use a Uniform prior over $(0, 1)$ which is the region which ensures a unique RE solution. This prior is combined with the likelihood function based on the two equations for π_t and x_t . We use a random walk Metropolis-Hastings algorithm to do posterior simulation using the first T of the artificially generated observations for $T = 10, 20, 50, 100, 1,000$ and $10,000$.

Table 2 presents posterior variances for various sample sizes. The results in Table 2 not only confirm the large sample theory derived in this paper, but also show that it is empirically useful. That is, in this RE model, it can clearly be seen that the posterior variance of ρ is going to zero much faster than for the unidentified parameters. Indeed the posterior for β is showing only slight divergences from its prior, even for very large sample sizes. But even the posterior variance for γ can be seen to be going to zero at very slow rate.

Table 2: Posterior Variances of Structural Parameters			
Number of observations	ρ	γ	β
T=10	0.027	0.059	0.079
T=20	0.014	0.054	0.079
T=50	0.007	0.050	0.079
T=100	0.004	0.043	0.079
T=1,000	0.001	0.019	0.076
T=10,000	9×10^{-5}	0.009	0.075

The posterior variance is not the only possible feature that the researcher could use as an indicator of identification. An alternative is simply to plot the posteriors for different choices of T . This is done in Figures 5, 6 and 7. In Figure 5, the posterior for ρ , clearly is converging in the manner implied by the asymptotic theory for identified models. However, in Figures 6 and 7, the posteriors for γ and β are converging much more slowly. The posterior for γ changes with T more rapidly than the posterior for β . However, there is clearly some updating of beliefs about β occurring (remember that the prior for this parameter is uniform).

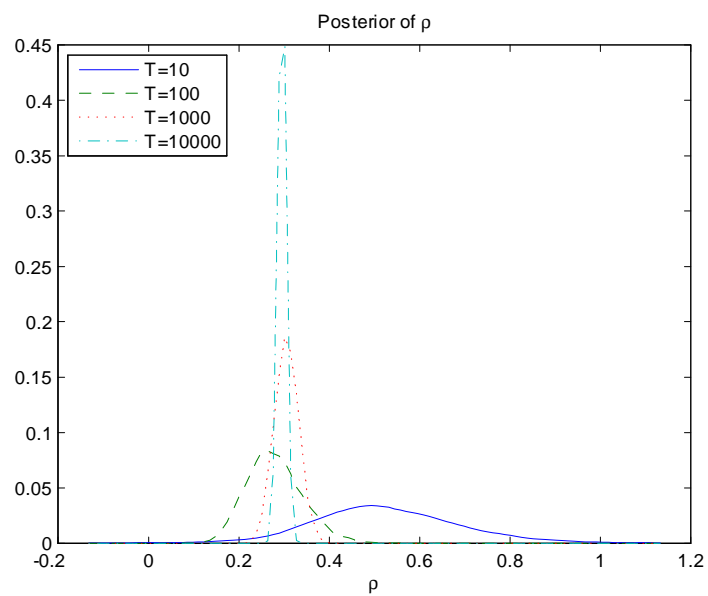


Figure 5

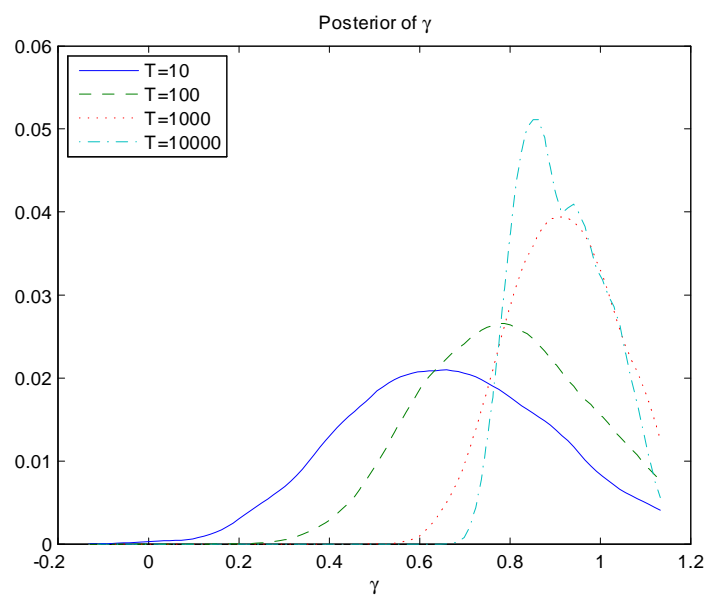


Figure 6

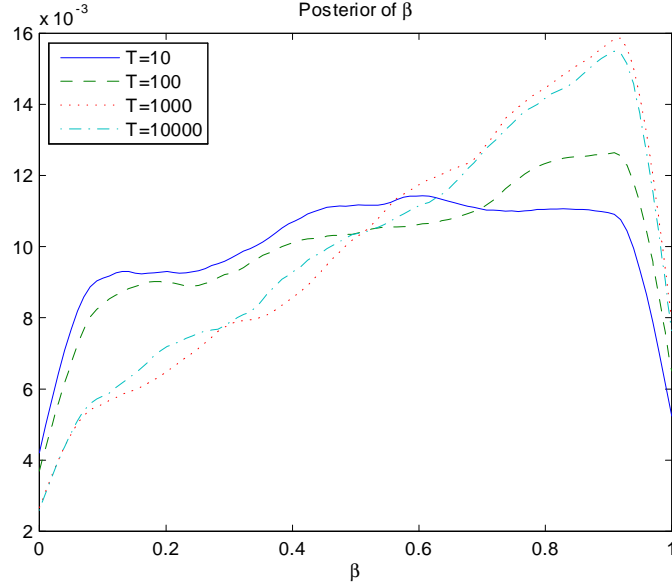


Figure 7

4.2.2 The NKPC with AR(2) Process for Output Gap

We now turn to the identified version of the NKPC, where we have an AR(2) process for the output gap. The structural parameters of this model (with no backward looking behavior) are $(\beta, \gamma, \rho_1, \rho_2)$ and the reduced form parameters are $(\alpha_1, \alpha_2, \rho_1, \rho_2)$. There is a one-to-one mapping between reduced form and structural form:

$$\alpha_1 = \frac{\gamma (\rho_1 + \beta_f \rho_2)}{1 - \beta_f (\rho_1 + \beta_f \rho_2)}$$

and

$$\alpha_2 = \frac{\gamma \rho_2}{1 - \beta_f (\rho_1 + \beta_f \rho_2)}.$$

Thus, all of the structural parameters are identified. However, if ρ_2 is near zero then identification will be weak. Accordingly, we use this example to investigate issues relating to the strength of identification.

The data generating process (DGP) is the same as in our previous example, except for its treatment of ρ_2 . We generate four different data sets with $\rho_2 = 0, 0.01, 0.1, 0.6$, respectively.¹⁵ The prior for (β, γ, ρ_1) is the same as that used

¹⁵In order to make these data sets as comparable as possible, we use the same seed for the random number generator for all DGPs.

in the AR(1) example. To this we add an independent $N(0.5, 0.1)$ prior for ρ_2 . Just as with the AR(1) example, this prior is combined with the likelihood function based on the two equations for π_t and x_t . We use a random walk Metropolis-Hastings algorithm to do posterior simulation using the first T of the artificially generated observations for $T = 10, 20, 50, 100, 1,000$ and $10,000$.

Table 3 reports results for the case where $\rho_2 = 0$. This artificial data set is the same as that used to produce Table 2. However, the model being estimated differs in this case. Since we are estimating an additional parameter, ρ_2 , it is not surprising that the posterior variances (especially for small sample sizes) are slightly larger in Table 2 than Table 3. However, the general pattern revealed by Table 3 is the same as Table 2. Table 4 (where $\rho_2 = 0.01$) also exhibits a similar pattern. These findings suggest that the Bayesian identification indicator based on large sample results will present useful information even if our estimating model is identified. Thus, it can be a useful indicator of weak identification.

In this model, the weakness of identification seems to impact mostly on β . That is, its posterior variance is decreasing very little over time in the cases where $\rho_2 = 0$ or 0.01 . With $\rho_2 = 0.1$ (see Table 5), we can begin to see clear signs that the posterior variance of β is decreasing with T . However, even in this case, the decrease of the posterior variance is quite slow. However, when $\rho_2 = 0.6$ (see Table 6), the posterior of all the parameters can be seen to be converging at roughly the same rate. This is reassuring since, in this case, all the parameters are strongly identified.

Table 3: Posterior Variances (DGP has $\rho_2 = 0$)				
Number of observations	ρ_1	ρ_2	γ	β
T=10	0.034	0.028	0.070	0.079
T=20	0.018	0.015	0.061	0.079
T=50	0.008	0.006	0.055	0.078
T=100	0.004	0.003	0.047	0.078
T=1,000	0.001	2×10^{-4}	0.020	0.078
T=10,000	8×10^{-5}	2×10^{-5}	0.011	0.076

Table 4: Posterior Variances (DGP has $\rho_2 = 0.01$)				
Number of observations	ρ_1	ρ_2	γ	β
T=10	0.034	0.027	0.066	0.078
T=20	0.018	0.015	0.063	0.078
T=50	0.007	0.007	0.056	0.078
T=100	0.004	0.003	0.046	0.078
T=1,000	0.001	2×10^{-4}	0.020	0.072
T=10,000	9×10^{-5}	2×10^{-5}	0.011	0.070

Table 5: Posterior Variances (DGP has $\rho_2 = 0.10$)				
Number of observations	ρ_1	ρ_2	γ	β
T=10	0.032	0.024	0.064	0.072
T=20	0.019	0.014	0.065	0.068
T=50	0.009	0.006	0.060	0.062
T=100	0.005	0.004	0.053	0.059
T=1,000	0.001	3×10^{-4}	0.030	0.048
T=10,000	8×10^{-5}	4×10^{-5}	0.013	0.036

Table 6: Posterior Variances (DGP has $\rho_2 = 0.60$)				
Number of observations	ρ_1	ρ_2	γ	β
T=10	0.037	0.028	0.070	0.047
T=20	0.015	0.012	0.057	0.022
T=50	0.006	0.006	0.043	0.014
T=100	0.004	0.004	0.033	0.010
T=1,000	0.001	0.001	0.012	0.003
T=10,000	6×10^{-5}	6×10^{-5}	0.001	4×10^{-4}

5 Concluding Remarks

This paper has examined the identification of the parameters of DSGE models, in the light of the widespread concern in the literature that the parameters may be either not identified or only weakly identified. In purely forward looking models, with no lags, the coefficients of the expectational variables are generically not identified since they do not enter the likelihood function. In forward looking models with lags, identification is dependent on the assumed structure of the dynamics, making it vulnerable to the Sims (1980) critique of ‘incredible’ identifying restrictions. In more complicated models with unobserved variables and no analytical solution, it is difficult to determine whether the models are identified. When the DSGE models are estimated by Bayesian methods, this lack of identification may not be evident since the posterior may differ from the prior even if the parameter is not identified and the posterior for unidentified parameters may also be updated as the sample size increases. These properties have been demonstrated both analytically and numerically, using familiar examples of unidentified or weakly identified rational expectations DSGE models.

We propose two Bayesian identification indicators. The first involves comparing the marginal posterior of a parameter with the posterior expectation of the prior for that parameter conditional on the other parameters. This can be computed as part of the MCMC estimation of a DSGE model using whatever real data set the researcher is working with. However, this indicator can be applied only in situations where parameters can be partitioned into a set that are known to be identified and another set for which identification is uncertain. This may not be possible when the researcher is working with the structural parameters of a DSGE model. Our second Bayesian indicator is more generally applicable and considers the rate at which the posterior precision gets updated as the sample size (T) is increased. For identified parameters the posterior pre-

cision rises with T , whilst for an unidentified parameter its posterior precision may be updated but its rate of update will be slower than T . This result assumes that the identified parameters are \sqrt{T} -consistent, but similar differential rates of updates for identified and unidentified parameters can be established in the case of super consistent estimators. This suggests a strategy where the researcher simulates larger and larger data sets and observes the behavior of the posterior as sample size increases.

We present an empirical illustration which shows the effectiveness of the first Bayesian identification indicator, in cases where it is applicable. Further empirical illustrations show the usefulness of our second Bayesian identification indicator, both for checking for the presence and the strength of identification.

References

- An, S. and F. Schorfheide (2007) Bayesian Analysis of DSGE models, *Econometric Reviews*, 26(2-4), 113-172.
- Andrle, M. (2010) A note on identification patterns in DSGE models, ECB Working Paper 1235.
- Bentai, L. (2010) Are policy counterfactuals based on structural VARs reliable?, ECB Working Paper 1188.
- Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer-Verlag: New York.
- Beyer, A. and R. Farmer (2004) On the indeterminacy of New-Keynesian economics. Working Paper Series 323, European Central Bank.
- Beyer, A., R. Farmer, J. Henry, and M. Marcellino (2007) Factor analysis in a model with rational expectations, NBER Working Paper 13404.
- Binder, M. and M.H. Pesaran (1995) Multivariate rational expectations models and macroeconometric modelling: A review and some new results. In M.H. Pesaran & M. Wickens (eds.), *Handbook of Applied Econometrics: Macroeconomics*, pp. 139-187. Oxford: Basil Blackwell.
- Binder, M. and M.H. Pesaran (1997) Multivariate linear rational expectations models: characterization of the nature of the solutions and their fully recursive computation, *Econometric Theory*, 13, 877-888.
- Canova, F. (2007) Comment on Bayesian analysis of DSGE models by S. An and F. Schorfheide, *Econometric Reviews*, 26, 113-172.
- Canova, F. and Sala, L. (2009) Back to square one: identification issues in DSGE models, *Journal of Monetary Economics*, 56, 431-449.
- Clarida, R., J. Gali and M. Gertler (1999) The science of monetary policy: a new Keynesian perspective, *Journal of Economic Literature*, 37, 1661-1707.
- Clarida, R., J. Gali, and M. Gertler (2000) Monetary policy rules and macroeconomic stability: Evidence and some theory, *Quarterly Journal of Economics*, 115, 147-180.
- Cochrane, J. (2007) Identification with Taylor Rules: a critical review, *NBER Working Paper* 13410.
- Dees, S., M.H. Pesaran, L.V. Smith and R.P. Smith (2009) Identification of New Keynesian Phillips Curves from a Global Perspective, *Journal of Money Credit and Banking*, 41(7).
- DeJong, D., B. Ingram and C. Whiteman (2000) A Bayesian approach to dynamic macroeconomics, *Journal of Econometrics*, 98, 203-223.
- Del Negro, M. and F. Schorfheide (2008) Forming priors for DSGE models (and how it affects the assessment of nominal rigidities), *Journal of Monetary Economics*, 55, 1191-1208.
- Drèze, J. (1976) Bayesian limited information analysis of the simultaneous equations model, *Econometrica*, 44, 1045-1075.
- Drèze, J. and J.-F. Richard (1983). Bayesian analysis of simultaneous equation systems, in Z. Griliches and M.D. Intrilligator, (eds.), *Handbook of Econometrics*, Vol. 1.
- Gali, J., M. Gertler and J. López-Salido (2005) Robustness of the estimates

of the hybrid new Keynesian Phillips curve, *Journal of Monetary Economics*, 52, 1107-1118.

Guerron-Quintana, P. (2010) What you match does matter: The effects of data on DSGE estimation, *Journal of Applied Econometrics*, 25, 774-804.

Hansen, L.P. and J.T. Heckman (1996) The empirical foundations of calibration, *Journal of Economic Perspectives*, 10(1), 87-104.

Heckman, J.J. (2010) Building bridges between structural and program evaluation approaches to evaluating policy, *NBER working paper 16110*.

Herbst E. (2010) Gradient and hessian-based MCMC for DSGE models, manuscript.

Hoogerheide, L., F. Kleibergen and H. van Dijk (2007) Natural conjugate priors for the instrumental variables regression model applied to the Angrist-Krueger data, *Journal of Econometrics*, 138, 63-103.

Iskrev, N. (2008) Evaluating the information matrix in linearized DSGE models, *Economics Letters*, 99, 607-610.

Iskrev, N. (2010a) Local identification in DSGE models, *Journal of Monetary Economics*, 57, 189-202.

Iskrev, N. (2010b) Evaluating the strength of identification in DSGE models. An a priori approach, manuscript.

Kadane, J.B. (1974) The role of identification in Bayesian theory. In S.E. Fienberg & A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics*, pp. 175-191. Amsterdam: North-Holland.

Kleibergen F. and S. Mavroeidis (2009) Weak instrument robust tests in GMM and the new Keynesian Phillips curve, *Journal of Business and Economic Statistics*, 27(3) 293-311.

Kleibergen F. and S. Mavroeidis (2010) Identification-pathology robust Bayesian analysis of DSGE models, manuscript.

Kleibergen, F. and H. van Dijk (1998) Bayesian simultaneous equations analysis using reduced rank structures, *Econometric Theory*, 14, 699-744.

Kleibergen, F. and E. Zivot (2003). Bayesian and classical approaches to instrumental variable regression, *Journal of Econometrics*, 114, 29-72.

Komunjer, I. and S. Ng (2010). Dynamic Identification of DSGE Models, manuscript.

Koop, G. and D. Poirier (1997) Learning about the cross-regime correlation in switching regression models, *Journal of Econometrics*, 78, 217-227.

Kydland, F. and E. Prescott (1996) The computational experiment: an econometric tool, *Journal of Economic Perspectives*, 10, 69-85.

Lindley, D.V. (1971) *Bayesian Statistics: A Review*, Philadelphia, SIAM.

Lubik, T. and F. Schorfheide (2004). Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review*, 94, 190-217.

Marschak, J. (1953) Economic measurements for policy and predictions, in W. Hood and T. Koopmans (eds) *Studies in Econometric Method*, 1-26, Wiley New York.

Mavroeidis, S. (2005) Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve, *Journal of Money, Credit, and Banking*, 37, 421-448.

- McCallum, B. (1979) On the observational inequivalence of classical and Keynesian models, *Journal of Political Economy*, 87, 395-402.
- Muller, U.K. (2010) Measuring prior sensitivity and prior informativeness in large Bayesian models, manuscript.
- Nason, J. and G.W. Smith (2008) Identifying the new Keynesian Phillips curve, *Journal of Applied Econometrics*, 23, 525-51.
- Pesaran, M.H. (1981) Identification of Rational Expectations Models, *Journal of Econometrics*, 16, 375-98.
- Pesaran, M.H. (1987) *The Limits to Rational Expectations*, Blackwell: Oxford.
- Poirier, D.J. (1998) Revising beliefs in unidentified models, *Econometric Theory* 14, 483-509.
- Pudney, S. (1982) The identification of rational expectations models under structural neutrality, *Journal of Economic Dynamics and Control*, 4, 117-121.
- Rothenberg, T. (1971) Identification in parametric models, *Econometrica*, 39, 577-591.
- Rubio-Ramirez, J., D. Waggoner and T. Zha (2010) Structural vector autoregressions: Theory of identification and algorithms for inference, *Review of Economic Studies*, forthcoming.
- Sargent T.J. (1976) The observational equivalence of natural and unnatural rate theories of macroeconomics, *Journal of Political Economy*, 84, 631-40.
- Sims, C. (1980) Macroeconomics and reality, *Econometrica*, 48, 1-48.
- Sims, C. (2002) Solving linear rational expectations models, *Computational Economics*, 20, 1-20.
- Smets, F. and R. Wouters (2003) An estimated dynamic stochastic general equilibrium model of the Euro area, *Journal of the European Economic Association*, 1, 1123-1175.
- Smets, F. and R. Wouters (2007) Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review*, 97, 588-606.
- Stock, J., J. Wright and M. Yogo. (2002) A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business and Economic Statistics*, 20, 518-529.
- Wallis, K.F. (1980) Econometric implications of the rational expectations hypothesis, *Econometrica*, 48, 49-73.